

What is claimed is:

1. A questionnaire analysis system comprising:  
means for inputting a questionnaire statement including free  
reply description in natural language;  
5 a network for transmitting a questionnaire reply statement,  
a database for accumulating said questionnaire reply  
statements transmitted through said network; and  
a text classification engine for reading out said questionnaire  
reply statements from said database and for learning a rule for  
10 classifying said questionnaire reply statement.
2. A questionnaire analysis system comprising:  
means for inputting a questionnaire statement including free  
reply description in natural language;  
a database for accumulating said questionnaire reply  
15 statement; and  
a text classification engine for reading out said questionnaire  
reply statement from said database and for learning a rule for  
classifying said questionnaire reply statement.
3. A questionnaire analysis system comprising:  
20 means for inputting a questionnaire statement including free  
reply description in natural language;  
a network for transmitting said questionnaire reply  
statement;  
a database for accumulating said questionnaire reply  
25 statement transmitted through said network;  
a text classification engine for reading out said questionnaire  
reply statement from said database and for learning a rule for  
classifying said questionnaire reply statement; and  
means for distributing said rule through said network

according to a request from a claimant.

4. The questionnaire analysis system according to claim 1, wherein said text classification engine includes:

morpheme analysis means for analyzing morphemes in all  
5 sentences in said questionnaire reply statement accumulated in  
said database;

category-text designating means for designating said category and text;

attribute selecting means for selecting attributes in plural  
10 questionnaire reply statements being read out from said  
database;

rule learning means for learning said rule for expressing said correspondence of text and category on the basis of said words selected by attributes by said attribute selecting means; and

15 rule output means for issuing said rule learned by said rule  
learning means.

5. The questionnaire analysis system according to claim 2, wherein said text classification engine includes:

morpheme analysis means for analyzing morphemes in all  
20 sentences in said questionnaire reply statement accumulated in  
said database;

category-text designating means for designating said category and text;

attribute selecting means for selecting attributes in plural  
25 questionnaire reply statements being read out from said  
database;

rule learning means for learning said rule for expressing said correspondence of text and category on the basis of said words selected by attributes by said attribute selecting means; and

rule output means for issuing said rule learned by said rule learning means.

6. The questionnaire analysis system according to claim 3, wherein said text classification engine includes:

5 morpheme analysis means for analyzing morphemes in all sentences in said questionnaire reply statement accumulated in said database;

category-text designating means for designating said category and text;

10 attribute selecting means for selecting attributes in plural questionnaire reply statements being read out from said database;

rule learning means for learning said rule for expressing said correspondence of text and category on the basis of said words  
15 selected by attributes by said attribute selecting means; and

rule output means for issuing said rule learned by said rule learning means.

7. The questionnaire analysis system according to claim 4, wherein said attribute selecting means computes a difference  
20  $\Delta SC(\omega)$  between a stochastic complexity (SC) of a test set without consideration of appearance of word and a stochastic complexity (SC) of a text set with consideration thereof, in each word  $\omega$  appearing in said text, and then selects said difference  $\Delta SC(\omega)$  as an attribute when said difference  $\Delta SC(\omega)$  is larger  
25 than said threshold  $\tau$ .

8. The questionnaire analysis system according to claim 5, wherein said attribute selecting means computes a difference  $\Delta SC(\omega)$  between a stochastic complexity (SC) of a test set without consideration of appearance of word and a stochastic

complexity (SC) of a text set with consideration thereof, in each word  $\omega$  appearing in said text, and then selects said difference  $\Delta SC(\omega)$  as an attribute when said difference  $\Delta SC(\omega)$  is larger than said threshold  $\tau$ .

- 5        9. The questionnaire analysis system according to claim 6, wherein said attribute selecting means computes a difference  $\Delta SC(\omega)$  between a stochastic complexity (SC) of a test set without consideration of appearance of word and a stochastic complexity (SC) of a text set with consideration thereof, in each  
10 word  $\omega$  appearing in said text, and then selects said difference  $\Delta SC(\omega)$  as an attribute when said difference  $\Delta SC(\omega)$  is larger than said threshold  $\tau$ .

10. The questionnaire analysis system according to claim 4, wherein said rule learning means:

- 15        forms said text set by replacing with an expression of  $(d_1, c_1)$ ,  $(d_2, c_2)$ , ...,  $(d_m, c_m)$  [where each  $d_i$  is a multi-valued discrete vector  $d_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{in})$  ( $i = 1, \dots, m$ ),  $\omega_{ij}$  is 1 when word obtained by attribute selection  $\omega_j$  ( $j = 1, \dots, n$ ) appears in said  $i$ -th text, or 0 otherwise,  $c_i$  expresses said value (label) of said  
20 category according to said  $i$ -th text and each  $c_i$  is 1 when belonging to a specific category, or 0 otherwise, and  $m$  is said number of texts];

25        selects said rules of if-then-else format and sequentially adds said selected rules to said stochastic decision list by employing said information quantity standard such as said extended stochastic complexity (SC) minimum principle or SC minimizing principle; and

      removes said rules one by one from said last one of said stochastic decision list, and clips continuously until none should

be removed from said viewpoint of said extended SC minimum principle.

11. The questionnaire analysis system according to claim 5, wherein said rule learning means:

- 5 forms said text set by replacing with an expression of  $(d_1, c_1), (d_2, c_2), \dots, (d_m, c_m)$  [where each  $d_i$  is a multi-valued discrete vector  $d_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{in})$  ( $i = 1, \dots, m$ ),  $\omega_{ij}$  is 1 when word obtained by attribute selection  $\omega_j$  ( $j = 1, \dots, n$ ) appears in said i-th text, or 0 otherwise,  $c_i$  expresses said value (label) of said category according to said i-th text and each  $c_i$  is 1 when  
10 belonging to a specific category, or 0 otherwise, and  $m$  is said number of texts];

- selects said rules of if-then-else format and sequentially adds said selected rules to said stochastic decision list by employing  
15 said information quantity standard such as said extended stochastic complexity (SC) minimum principle or SC minimizing principle; and

- removes said rules one by one from said last one of said stochastic decision list, and clips continuously until none should  
20 be removed from said viewpoint of said extended SC minimum principle.

12. The questionnaire analysis system according to claim 6, wherein said rule learning means:

- forms said text set by replacing with an expression of  $(d_1, c_1),$   
25  $(d_2, c_2), \dots, (d_m, c_m)$  [where each  $d_i$  is a multi-valued discrete vector  $d_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{in})$  ( $i = 1, \dots, m$ ),  $\omega_{ij}$  is 1 when word obtained by attribute selection  $\omega_j$  ( $j = 1, \dots, n$ ) appears in said i-th text, or 0 otherwise,  $c_i$  expresses said value (label) of said category according to said i-th text and each  $c_i$  is 1 when

belonging to a specific category, or 0 otherwise, and m is said number of texts];

selects said rules of if-then-else format and sequentially adds said selected rules to said stochastic decision list by employing  
5 said information quantity standard such as said extended stochastic complexity (SC) minimum principle or SC minimizing principle; and

removes said rules one by one from said last one of said stochastic decision list, and clips continuously until none should  
10 be removed from said viewpoint of said extended SC minimum principle.

13. A computer program product for analyzing questionnaire reply which comprises:

a morpheme analysis procedure for analyzing morphemes in  
15 all sentences in said questionnaire reply statements accumulated in a database;

a category-text designating procedure for designating said category and text in said text classification engine;

an attribute selecting procedure for selecting attributes in  
20 plural questionnaire reply statements being read out from said database;

a rule learning means for learning said rule for expressing said correspondence of text and category on said basis of said words selected by attributes by said attribute selecting  
25 procedute; and

a rule output procedure for issuing said rule learned by said rule learning procedure.